Benchmarking Big Data Research Project

Report from the University of East Anglia

ESRC Business and Local Government Data Research Centre

University of East Anglia

Document Information

Status: Final

Authors: James Cornford and Catherine Levin

Acknowledgement

The ability to undertake this vital piece of research was made possible by the contributions from the following Fire and Rescue Services. This is the first time that CFOA has utilised the crowd funding methodology and this enabled the financial opportunity to be spread across a number of contributors. The success of this methodology means that we are likely to make use of this process to fund further research for the benefit of UK FRS, our firefighters and our communities.

- Hampshire
- Cheshire
- Buckinghamshire
- Dorset & Wiltshire
- London
- Ordnance Survey
- Tyne and Wear
- Leicestershire
- Cornwall
- East Sussex
- Humberside
- Gloucestershire
- Lancashire
- South Wales
- Durham and Darlington
- Devon and Somerset

The authors would also like to thank all the individuals who gave up their time to be interviewed or helped in other ways with this project.

Contents

Figures and Tables

Executive Summary

Background

This document describes and creates a baseline for the Fire and Rescue Service's (FRS) handling of big data sets shared by other public services.

The FRS have transformed over recent decades from a mainly reactive 'brigade' to a primarily proactive and preventative service. This transformation has involved the profiling or stratification of fire risk at a household level in order to target preventative activity more efficiently and effectively on vulnerable groups. This stratification is reliant on access to timely and comprehensive data.

The SAfER project has pioneered the sharing by the NHS of a subset of the 'Exeter data' (the date of birth and address of all individuals registered with a GP and over the age of 65). This initiative, a first example of national level big data sharing with the FRS, is used in this report as a litmus test of the challenges involved in making use of big data.

Methodology

This study use a big data value chain approach, which explores the issues involved in the acquisition and storage, preparation and cleansing, linkage and integration, analysis, modelling and visualisation, decision-making and evaluation of big data.

The value chain was used to structure telephone interviews with data analysts and other data specialists in 36 of the 45 FRS in England. The interviews were supplemented with context interviews and documentary sources.

Findings

Thanks to the work undertaken by the SAfER project, there were few issues involved in the acquisition of the Exeter data. However, negotiating access to other data sets, at a national or local level has often been a time-consuming and frustrating business.

The preparation and cleansing of large data sets to remove or correct anomalies, and miscoding and to prepare the data for linkage or layering represents a major, time-consuming task in most cases. For the Exeter Data, the lack of a Unique Property Reference Number (UPRN) was a particularly significant issue.

The data was seen as most valuable when combined – linked or layered - with other data sets to create a more comprehensive picture of individual households. For example linking the age of household members with household composition (e.g., single person households) and indicators of frailty or incapacity.

Few analysts reported major issues with analysing large volumes of data as they were generally using familiar risk stratification algorithms of varying sophistication.

Most analysts reported that the analyses of the Exeter data, generally in conjunction with other sources, were being used in the operational targeting of resources, mainly in the form of preventative Safe and Well visits (formerly Home Fire Safety Visits).

There was some anecdotal evidence that the additional Exeter data was helping to improve the targeting of preventative work on the most vulnerable individuals and households.

Most respondents found the main challenges to their use of data related to a lack of time, rather than lack of skills or a lack of technical facilities.

There has been limited formal evaluation of the data by individual FRS so far (although an evaluation of the SAfER project is planned).

The organisational "big data maturity" of FRS, as measured by both self-assessment and by progress along the big data value chain, suggests a number of leading organisations, a larger number of strongly performing organisations and a smaller number of lagging services.

Recommendations

The major barrier to increasing use of data in decision making, from the analysts point of view, is considerable time resource required for data preparation and cleansing. If data shared by other public services can be centrally "pre-cleansed" before it is released to individual FRS it is more likely to be used or will be used more quickly and widely.

FRS need to 'Close the Feedback Loop' by systematically reviewing the value of data over the short, medium and longer term. The development of a standard methodology or tool kit for such evaluation would be helpful to individual FRS.

The community of data analysts within the FRS is fragmented and some data analysts feel isolated. More efforts could be made to bring the analyst community together and to support the development of individual competencies and the collective knowledge-base.

Research revealed a lot of 'tactical' knowledge – tips and tricks that analysts use to acquire, prepare, analyse and visualise data. Analysts could benefit greatly from the sharing of this, largely informal, knowledge.

Further Research

The development of a comprehensive 'Maturity Capability Model,' drawing on models developed elsewhere and more intensive work with a small number of leading FRS, could create a more effective tool for FRS to benchmark themselves.

Progress towards becoming a fully data-led organisation represents a complex organisational change, requiring effective leadership and culture change as well as operational and policy innovation. Further, in depth, research into the questions of organisational culture and organisational leadership would also provide valuable lessons.

Introduction and background

This is the final report of the project Benchmarking Big Data in the Fire and Rescue Service, carried out by the University of East Anglia (UEA) and Elginfire Consulting Ltd for the Chief Fire Officers Association (CFOA).

This study has been sparked by the growing interest in the opportunities to use (or re-use) administrative data across public services to support improved outcomes as a result of, for example, better targeting of resources, modelling to support early intervention and preventative work, and the personalisation of services for clients and customers.

However, such systematic sharing of data presents a number of challenges – technical, organisational, legal, ethical – and has created new demands for skills and competences, ranging from information management and information governance to data analytics and data visualisation.

We know relatively little about the challenges and implications of systematic sharing of big data (as opposed to case-by-case information sharing). In particular, there is little evidence that would enable organisations, or whole sectors, to create empirically-based benchmarks or identify good or promising practices in this field.

In 2014, the Fire Services Research and Training Trust awarded CFOA a grant of £57,460 for a project called Sustained Action for Elderly Risk (SAfER). This project has sought to use a subset of fields from the National Health Service's Exeter System (also known as the NHAIS[1]) to create a list including the address, year of birth and gender of each individual over 65 who is registered with a GP in England and Wales.

This transfer of data has been intended "to ensure that FRSs target preventative resources more effectively, at a time where the ageing demographic means fire deaths and injuries will increase significantly for the first time in 30 years."[2]The SAfER project has its own, separate, evaluation process and this report is not intended to replicate that work but rather to explore the wider question of the capacity of the FRS to handle big data.

Each FRS in England received a subset of the Exeter Data in 2015 (although Cheshire FRS had pioneered the use of Exeter data from 2008).[3] While there has been some co-ordination, for example through the SAfER project and the CFOA Integrated Data and Research Programme (IDRP), each FRS has been free to use the data as they see fit (within the applicable Information Governance framework).

[1] National Health Applications and Infrastructure Services
[2] Evan Morris presentation of the SAfER project [http://cfoaservices.co.uk/attachment/download/link/id/112/]
[3] Centre of Excellence for Information Sharing (n.d.) *Cheshire Fire and Rescue's Innovative use of the Exeter Health Data* [available at http://informationsharing.org.uk/wp-content/uploads/2015/11/Exeter-health-data-Cheshire-Fire-Case-Study-1.pdf]

The SAfER project represents the first system wide experience of the FRS using big data from other public services. The experience of the Exeter Data is being used here as a good 'litmus' test of the level of maturity of each FRS with regards to the use of such large data sets.

The sharing of the Exeter Data can be seen as part of a wider, developing relationship between the FRS and the health sector. In 2015, a new partnership was established by the FRS and the NHS, with Public Health England, CFOA, the Local Government Association and Age UK, to recognise 'Fire as a Health Asset.'[4]

Over the winter of 2015-6, a pilot scheme was run to explore FRS interventions to reduce the risk of winter-related ill health in vulnerable groups of people, mainly using FRS Safe and Well visits to identify households at risk of falls, social isolation, cold homes and flu.[5]

In 2016, the partnership published a joint consensus statement[6] in which they pledged to offer 'an integrated approach to targeting through the better co-ordination, prevention and early intervention that has been demonstrated to increase the reach and impact of all services.'[7] The partnership has also jointly designed principles for FRS Safe and Well visits to enhance their value for the health service and other agencies.[8]

The sharing of the Exeter Data provides a unique opportunity to understand, in a systematic way, how such big data initiatives based on sharing data across different public sector 'silos' can work, the challenges they face and the new and promising practices that have been developed to overcome those challenges.

Further, the systematic and system-wide sharing of such data means that we are able to create a crude national benchmark for the FRS community which can be used by individual FRS to assess progress and which can provide CFOA with a baseline against which to measure progress and development of the field as a whole.

[4] https://www.england.nhs.uk/2015/08/frs-partnership/ and https://www.england.nhs.uk/ourwork/ltc-op-eolc/older-people/fire/.

[5] See Public Health England (2016*) Evaluation of the impact of Fire and Rescue Service interventions in reducing the risk of harm to vulnerable groups of people from winter-related illnesses*. London: PHE.

[6] https://www.england.nhs.uk/wp-content/uploads/2015/09/joint-consens-statmnt.pdf

[7] Ibid., p.2.

[8] https://www.england.nhs.uk/wp-content/uploads/2015/09/safe-well-visit-pinciples.pdf

The Study

The study design used a big data value chain model (see Figure 1) to structure the approach. This simple, linear model breaks down the use of big data into a number of stages. At the highest level, the value chain is divided into three stages – data discovery, data integration and data exploitation, some of which are further broken down. For this study, we have combined some stages and added an evaluation and development stage, a significant omission from the original model.

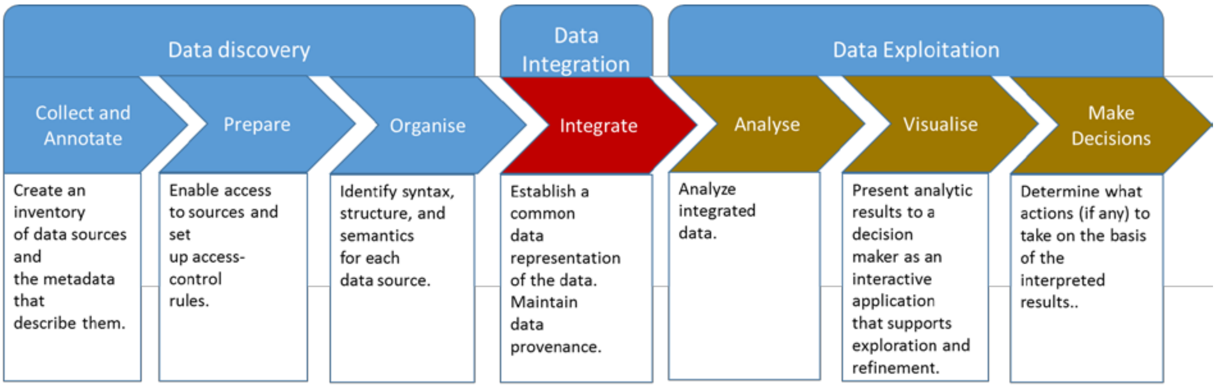

Figure 1: 'Big Data Value Chain' Source; Miller and Mork (2013: 58)

A big advantage of the Exeter Data, in terms of benchmarking FRS big data maturity, has been that access to the data, including information governance arrangements, was centrally negotiated. This means that the earlier stages of data discovery were effectively undertaken by the SAfER project for the FRS community as a whole, providing a more or less common starting point. However, on receipt of the Exeter Data each FRS went through largely independent processes of preparing, organising, integrating, analysing and visualising the data and making decisions based on the data.

The big data value chain approach was used to design a telephone questionnaire that served as the main data collection device (see appendix 2). The questionnaire comprised eight sections:

A. About the Informant
B. Acquisition and Storage
C. Data Preparation and Cleansing
D. Data Linkage and Integration
E. Data Analysis, Modelling and Visualisation
F. Use of the Exeter Data in Decision-making (planning, targeting, operational decision-making)
G. Evaluation and Development
H. Final Questions (further involvement in research)

The survey instrument was piloted with a small number of informants to ensure that the questions made sense and that the survey could be undertaken in approximately 30-40 minutes. Some small adjustments were made after the pilot stage.

The telephone survey was complemented by two key informant interviews, also conducted over the telephone using a topic guide (Appendix 4), and a systematic review of the relevant literature (see appendix 1 for the literature consulted).

There are 45 FRS in England. A named respondent was contacted in each of the 43 mainland FRS[9]. If there was no response to the first contact a second attempt was made and on some occasions a third contact was required. Responses to the telephone survey were received from 38 FRS. Two services declined to be interviewed but did provide information on the use of the Exeter Data so the total number of telephone interviews was 36 – an 80% response rate. Collectively, respondents' organisations had received more than 9 million Exeter Data records.

Where possible and where it is of interest we have reported based on the FRS 'families' – Metropolitan (here including London), Combined and County Services. The response rates for each family suggest that we have effective coverage of all three 'families' (See Table 1).

Family	Total Number	Responded (response rate)	Completed Interview (response rate)
Metropolitan	7	7 (100%)	7 (100%)
Combined	22	19 (86%)	17 (77%)
County	14	12 (86%)	12 (86%)

Table 1 Response rate by FRS 'Family'

Responses to the closed questions in the telephone survey were coded in an Excel spreadsheet and simple descriptive statistics were created. Narrative answers and the key informant interviews were transcribed and, together with the secondary literature, were imported to QSR NViVo, a qualitative analysis package.

[9] The Isle of Wight and Isles of Scilly were excluded from this study.

Key Findings

The respondents and their experience and self-assessed skills

Respondents had a range of job titles. The most commonly used terms apart from manager were risk, business, information, intelligence and analyst. Figure 2 shows a word cloud of the respondents' job titles. Most respondents were technical analysts mainly supporting operational aspects of FRS.

Figure 2 Word cloud based on the frequency of terms in respondents' job titles

Most respondents had significant experience of working in an FRS (see figure 2) and there was little variation in this pattern across the three FRS families.

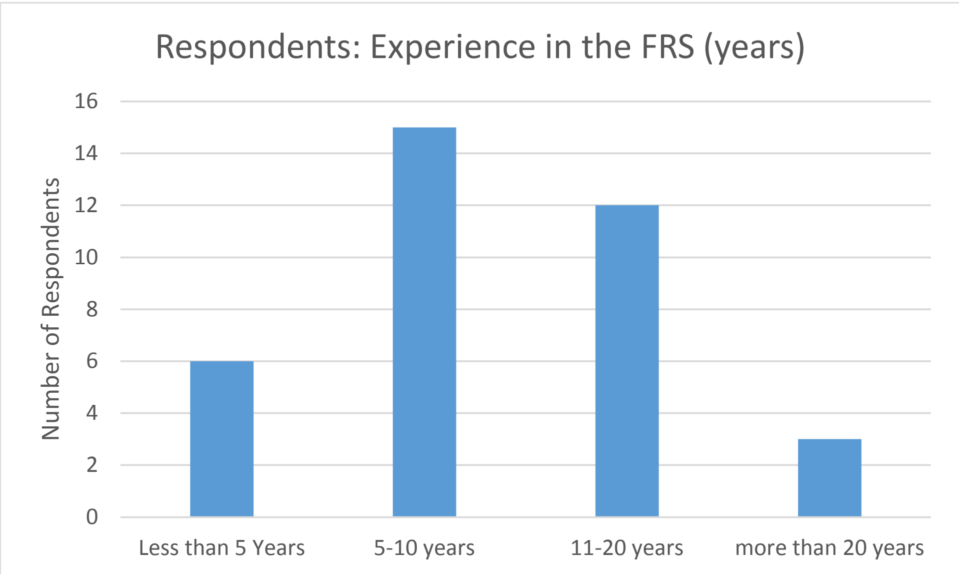

Figure 3: Respondents' experience in the FRS.

The respondents that we talked to were generally highly positive about their own skills in relation to information management - data handling and governance. On average they rated themselves at 7.7 out of 10 (where 10 is world class). Two respondents rated themselves as a 10/10 on this question and 7 rated themselves as 9/10. There was little systematic variation by FRS family but a slightly higher self-evaluation from the Combined services.

Respondents were slightly less confident about their analytics – that is, statistical analysis and modelling – skills. The average here was 7.4 and the highest self-assessment was a 9. No individual rated themselves below 5 in relation to either set of skills. Again, there was a similar pattern in each family and, again, the Combined services had a slightly higher reported self-evaluation.

Respondents were less confident about the maturity of their organisation in working with large data sets. On average, they rated their organisation as 6.1 out of 10 (where 10 was "world class") at the start of the interview. While some FRS were confident that they were close to world leading in their capacity there was more variation in this self-assessment task with some services rating themselves as low as 3 or 4 (see figure 3). In terms of the FRS families, the County FRS respondents gave the lowest average self-assessments (5.9), but the Combined (6.1) and Metropolitan (6.3) FRS averages were not much higher.

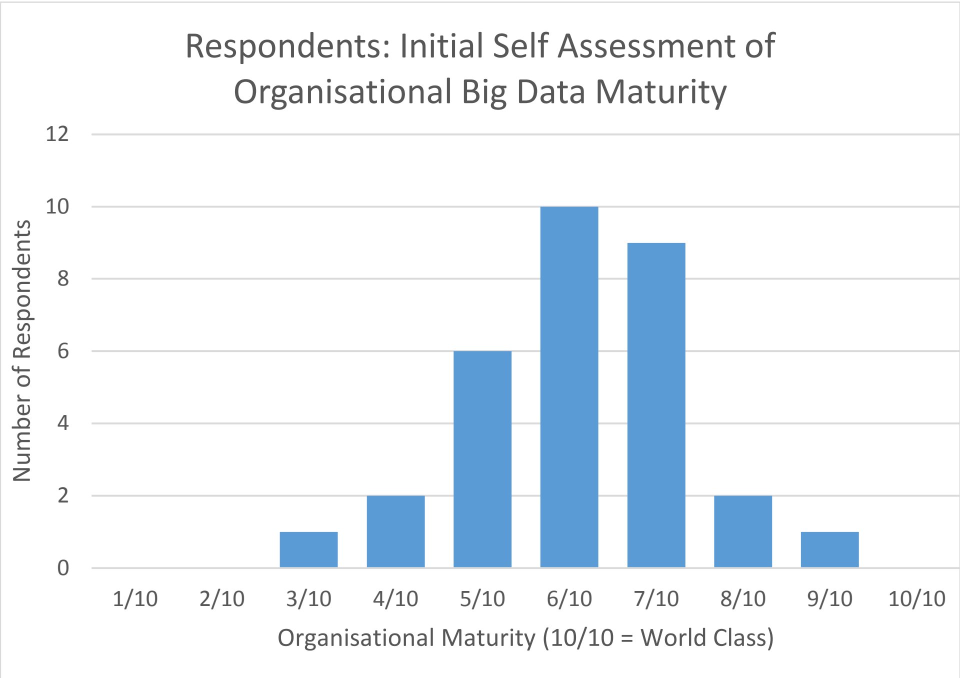

Figure 4 Respondents Self-assessment of Organisational Big Data maturity

Acquisition and Storage

Respondents represent organisations that received over 9 Million Exeter Data records.[10] The volumes of data delivered varied significantly. The London Fire Brigade received around 1,700,000 records while Warwickshire FRS received just 46,000 records. The Exeter Data was delivered via a secure email as an Excel spreadsheet and was stored on a local server, PC or laptop in all cases (and also stored within the cloud in one case and in Civica CFRMIS in another case).

Storage	**% of respondents**
Civica CFRMIS	2.78
Other risk management software	-
Local server PC or lap top	97.22
Cloud storage	2.78
Removable storage	-
Other	-

Table 2: Exeter Data Storage Solutions

Access to the Exeter Data appears to be restricted to the Research and Analysis Specialists with each FRS. In a few cases, senior board members also had access.

Access	**% of respondents**
Research and Analysis specialists	100.00
Senior Management Board members	5.56
Other	2.78
Chief Fire Officer	-
External partners (e.g., Universities or Consultants)	-

Table 3 Who has access to Exeter data

The Exeter Data is covered by an Information Sharing Agreement that is signed by each of the FRS in receipt of the data. Of the respondents, 77.78% had read the agreement themselves while others, who had not read the Agreement were aware of its existence and were confident that others in the organisation had read the agreement and that their usage was compliant.

Overall, acquisition and storage of the Exeter Data provided little challenge to the FRS with just 16.67% of respondents reporting problems at this stage in the value chain.

[10] Each record is an individual aged over 65 registered with the NHS. According to the ONS (2015) there were 9,537,708 individuals aged 65 or over in England in 2014 so we have approximately 95% population coverage.

Data Preparation and Cleansing

Almost all respondents (92.11%) reported that their organisation had sought to review, document, prepare and cleanse the Exeter Data.

A number of common issues arose at this point in the value chain. The most commonly reported concerns were the lack of a UPRN (Unique Property Reference Number) and the presence of postcodes from outside the organisation's area (and the absence of data from inside the area).

Respondents reported a range of strategies for overcoming the lack of a consistent UPRN and reconciling address data more generally. Software is available to automate the matching of most addresses to UPRNs, but there is usually still significant manual matching required.

Strategies for filling in the gap caused by misallocated data vary from simply accepting the missing data to attempts to contact neighbouring organisations and arranging swaps (although some respondents were waiting to receive legal or information governance advice on the appropriateness of these swaps).

Other data cleansing issues have included identifying and removing concentrations of older individuals, such as hospitals, care homes and sheltered housing, which are dealt with by other processes.

Data Cleansing issue	**% of respondents reporting**
Lack of UPRN	93.33
Postcodes from outside your area	85.29
Missing postcodes	73.53
Other	62.86
Spelling mistakes/typos	37.14
Implausible dates of birth	20.59

Table 4 Data cleansing issues

Common Data Management Challenges	**% of respondents reporting**
Reconciling address data	64.71
Other issues	63.64
Time to carry out the work	62.86
Strategies for filling the gaps in the data	42.42
The limitations imposed by the ISA	3.03

Table 5 Common Data Management Challenges

The main challenges associated with data cleansing have been about finding resources for what is often a time consuming activity. Most FRS struggled with this task in relative isolation (although there has been some contact to try to correct geographical misallocation). There may be opportunities to deal with some data cleansing and other data quality issues centrally or jointly. We understand that the lack of UPRN on Exeter Data will be addressed centrally in the future releases, but there are important lessons for other data sets and perhaps for other data quality issues with regards to Exeter Data.

For some respondents, the challenge of data cleansing, in particular dealing with the geocoding of the data, represented an effective barrier to using the data or using all of the data.

Data Linkage, Layering and Integration

Most respondents argued that the Exeter Data was much more useful when used in conjunction with other data. This requires the data to be linked or layered with other datasets. The vast majority of organisations (86.49%) had linked or layered the Exeter Data with other data sets. Around two thirds of respondents indicated that they had plans to undertake further linkage, layering or integration of the Exeter data.

In many cases (20), this included respondents who had already undertaken some work in this area and sought to extend it. In a small number of cases (2) it was organisations that had not yet worked with the Exeter Data in this way, but had plans to do this work.

A few organisations (12) had completed their integration of Exeter Data and had no further plans to work with the current release of the data. Just one organisation that responded to both questions had neither undertaken integration work, nor had plans to do this in the future, in this case because the data was not seen as adding to existing resources.

Common data sets that were linked or layered with the Exeter Data include commercial products such as Experian's MOSAIC and, much less commonly, the Caci Acorn classification. The list also included data from the organisation's Incident Recording Systems, other GIS systems and other geographic and demographic data sets. A few organisations had plans for further linkage with these, relatively established data sets but respondents reported a wide range of plans to link to further new data sets.

The main challenge that respondents reported in linkage and layering of data concerned available resources (including the time of skilled individuals - 52.17%) and, much less significantly, a lack of suitable technical infrastructure (9.09%).

Other data sets used in linkage and layering	% of Respondents
Other	84.38
IRS/Incident recording system	75.00
Experian MOSAIC	71.88
Other GIS (Geographical Information System)	50.00
NLPG/Geoplace Data	31.25
Census Data	18.52
Caci ACORN	6.25

Table 6 Other data sets used in linkage and layering

The provision of Exeter Data set stimulated a considerable appetite for data among the analyst community. In particular, many respondents expressed interest in the possibility of accessing other administrative data sets. Appendix 4 contains a summary of the various sources mentioned and the FRS that mentioned them. One of the most positive outcomes of the provision of Exeter data has been the demonstration effect and the stimulation of creative thinking about how data can be used to plan, target and evaluate services.

Perhaps the most important shift that the Exeter Data has stimulated in many FRS is a shift from spatial focus, often effectively at the level of postcodes if not postcode areas, towards a focus on individuals linked to specific residences. This is more than just a refinement of scale, representing a significant shift of focus within the FRS from buildings (potentially containing people), to people (potentially exhibiting specific vulnerabilities and behaviours).

Data Analysis, Modelling and Visualisation

Roughly four-fifths (81.1%) of organisations had undertaken some analysis, modelling or visualisation using the Exeter Data (either on its own or linked to other data sets). A little over half of respondents (55.2%) indicated that they would be undertaking further analysis in the future. A smaller proportion (37.9%) had no plans and a few (6.9%) were unsure.

Unsurprisingly, most of the analysis used the data in existing risk stratification models, which are then used, or will be used, to prioritise preventative work such as Safe and Well visits.

In some cases, analysts are focusing on a more elderly subset of the Exeter Data, for example over 80s or over 85s, or on specific defined geographical areas.

The main challenges confronted were again dominated by a lack of resources (65.38% of those who responded) and, to a lesser extent, problems with a lack of technical infrastructure (24.0%). Only one organisation saw a lack of data management and data analysis skills as a significant challenge.

Most FRS were working relatively independently on the analysis of data, including, Exeter Data but a significant minority were working with external organisations including other FRS, universities, software vendors and consultants.

Partners in working on Exeter Data	% of Respondents to this question
Other (please specify)	40.00
Other FRS	23.33
Universities	20.00
Software vendors	13.33
Consultants	10.00

Table 7 FRS Partners in working on Exeter data

Use of Data in Decision-making (planning, targeting, operational decision-making)

The ultimate value of data, it can be argued, resides in its use to inform operational and strategic decision-making. However, such use is only possible after the rest of the value chain has been negotiated. Impressively, four fifths (80%) of respondents could identify ways in which their organisation had made use of analysis, models or visualisation that draw on the Exeter data to support strategic or operational decisions and around half (52.17%) of respondents indicated that they had (further) plans to use analyses of the Exeter Data in decision-making.

All organisations that had used analyses of Exeter Data had used it, as intended, in targeting and risk stratification models. However, most also reported using or planning to use, the data in resource allocation decisions.

Even more impressively, **the vast majority (81.59%) of those who had used the Exeter Data had, or expected to have, some evidence of the impact of these decisions on FRS outcomes or Key Performance Indicators.**

Evaluation and Development

In spite of the expectation that data-driven decisions were having a positive impact, many FRS had not yet considered the formal evaluation of the Exeter Data. In part, this reflects their expectation that the SAfER project will be undertaking a systematic evaluation and therefore evaluation of the data may have been deemed unnecessary or wasteful duplication. Informal feedback from crews undertaking Safe and Well visits did provide some anecdotal evidence of improved targeting. It may be too early in the exploitation of the Exeter Data to expect formal systemic evaluation of the impact of the data, but evaluation plans could be in place and baseline data gathered.

In terms of development of big data capacity, respondents were generally enthusiastic about receiving support including mentoring from other FRS, expert guidance in data-driven decision-making and formal training in data management and analysis.

Forms of Support	**% of respondents**
Mentoring from other Fire and Rescue Services	84.85
Training in data management and/or analysis techniques	78.79
Expert guidance on the use of data in decision making from outside the FRS	78.79
Specialist software for data management and analysis	51.52
Other (please specify)	42.42

Table 8 Forms of Support

Big Data Maturity

The concept of organisational maturity has resulted in the development of formal maturity models in a number of fields over the last three decades. Perhaps the best known application of a maturity model is the Capability Maturity Model for Software developed by Carnegie Mellon university for the US Department of Defense,[11] but has also been applied to other fields (for example, project management).[12] Where they have been developed, such models are widely used for benchmarking. There has been some work on the development of 'Big Data' Capability Maturity Models, for example by companies such as IBM and CSC,[13] but an established cross sector model has yet to emerge.

Typically, maturity models focus on repeated operations. A simple, but robust, generic capability maturity model might have four levels (see figure 4 below) ranging from non-existent or random processes, through replicable to improvable and ultimately optimisable levels. Models that are more sophisticated can be developed but they are likely to require large volumes of systematically collected data over a number of time periods.

- Level 4: Optimisable - highly replicable success, exceeds expectations, fully mapped processes, excellent documentation, learning from internal and external sources
- Level 3: Improvable - increasingly replicable success, improvable processes, good documentation, some learning from internal and external sources…
- Level 2: Replicable - generally replicable success, some process and documentation, conformance orientation, copying external sources…
- Level 1: Random - some success but more by luck than judgement; little process, little documentation…

Figure 5 Generic Four Level Maturity Model

Because we only have a single big data case – the Exeter Data – to examine, we cannot take this approach to modelling maturity in the FRS yet. We have therefore had to adopt an alternative approach to benchmarking.

One alternative source of data on organisational maturity is a simple self-assessment. At the end of the interview, we asked respondents for a second self-assessment of organisational maturity, to see if reflecting on the Exeter Data experience would allow respondents to reassess their organisation's maturity

[11] See Paulk et al. 1993 and http://cmmiinstitute.com/.
[12] See e.g., the Project management Institute's OPM3 – Organisational Project Management Maturity Model (https://www.pmi.org/business-solutions/assessment-benchmarking/organizational) and the UK Office of Government Commerce-developed Portfolio, Programme, and Project Management Maturity Model (https://www.axelos.com/best-practice-solutions/p3m3).
[13] See e.g., http://www.csc.com/big_data/insights/111325-big_data_maturity_model_planning_your_strategy and http://www.ibmbigdatahub.com/blog/maturity-model-big-data-and-analytics

(see Figure 4 and compare with Figure 3). There was little change overall and where there were changes, respondents tended increase their assessments of their organisation’s big data maturity.

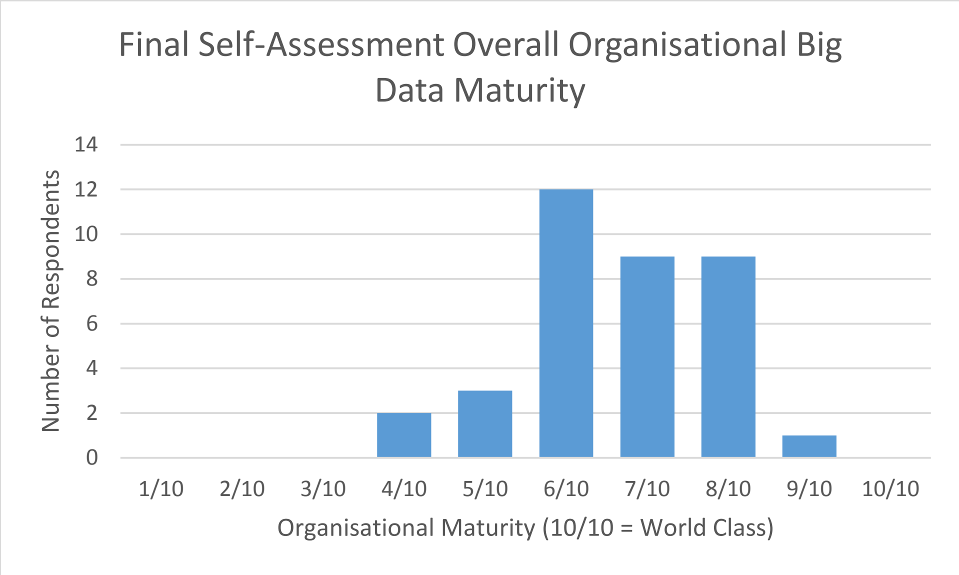

Figure 6 Self-Assessment Overall Big Data Organisational Maturity

We have also sought to create a more objective model of maturity using the big data value chain to benchmark organisations in terms of their progress along the chain. Using the responses to the questions detailed in table 9 we have been able to group the FRS into a number of maturity levels. A simple procedure of adding the number of times the respondent was able to answer “yes” to these questions (or give details for Question G1) gives us a rough indication of the degree of progress made. Figure 6 below shows the full distribution of respondents. On this basis, out of the 38 responses:

- 10 (26.3%) were able to respond positively to either 9 or all 10 of the questions – the most mature organisations;
- a larger group of 17 (44.7%) of the respondents were able to answer 6, 7 or 8 questions affirmatively - we might see these as well on their way to maturity;
- seven (18.4%) respondents were able to answer 3, 4 or 5 questions in the affirmative – indicating organisations that have made some initial progress along the value chain; and,
- just 4 respondents answered fewer than 3 of the ten questions positively – indicating very little development.

Table 10 below shows the distribution of maturity by FRS Family.

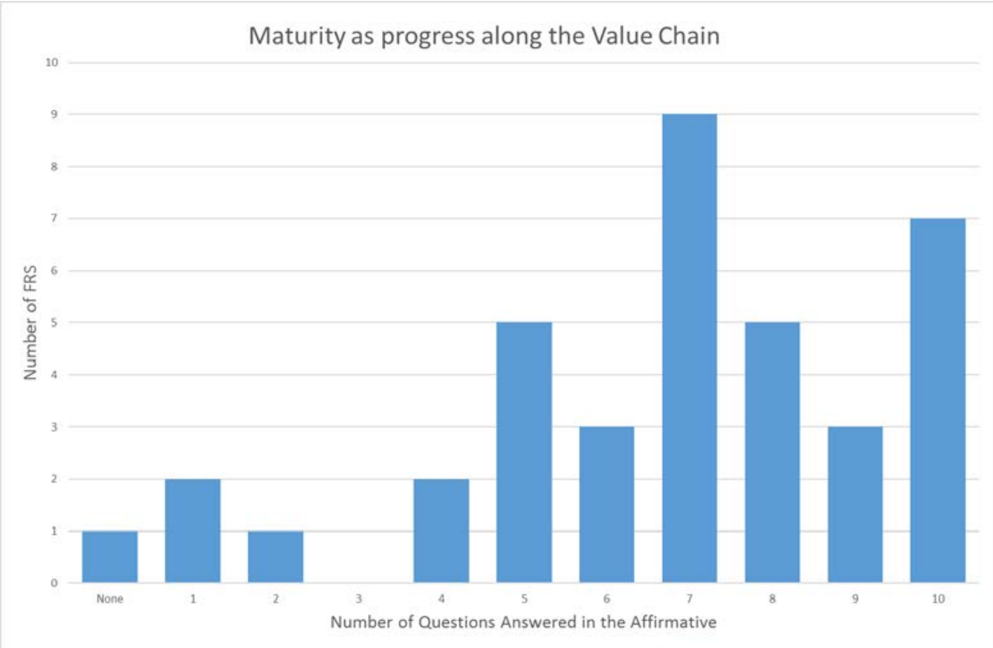

Figure 7 Maturity as Progress Along the Value Chain

Question	Rationale for inclusion
B0 How many Exeter Records did you receive?	Demonstrates basic awareness of data
B3 Have you seen the Information Sharing Agreement which covers the Exeter data set?	Demonstrates awareness of Information Governance
C1 Has your organisation sought to review, document, prepare and cleanse the Exeter data?	Indicates data management capability
D1 Has your organisation linked the Exeter data to, or layered it with, existing data sets or products?	Indicates core data handling capability
D3 Does your organisation *plan* to link the Exeter data to other (further) data sets or products and if so which ones?	Indicates core data handling capability
E1 Has your organisation undertaken any analysis, modelling or visualisation using the Exeter data (either on its own or linked other data sets)?	Indicates core data analysis capability
E3 Does your organisation *plan* to undertaken any (further) analysis, modelling or visualisation using the Exeter data (either on its own or linked other data sets)?	Indicates core data analysis capability
F1 Has your organisation made use of analysis, models or visualisation that draw on the Exeter data to support strategic or operational decisions?	Indicates ability to apply analysis to decision making
F3 Do you have or expect to have evidence of the impact of these decisions on FRS outcomes or KPIs?	Indicates ability to effectively apply analysis to decision making
F5 Does your organisation plan to make use of analyses, models or visualisation of the Exeter data to support strategic or operational decision-making?	Indicates ability to apply analysis to decision making
G1 How does your organisation plan to evaluate the benefits of the Exeter data?	Indicates continuous improvement and learning capability

Table 9 Questions used to create Data Maturity Model

Maturity Level Family	1 (Lowest)	2	3	4 (highest)	Totals
Metropolitan	0	1 (14.3%)	3 (42.9%)	3 (42.9%)	7
Combined	3 (15.8%)	1 (5.3%)	9 (47.4%)	6 (31.6%)	19
County	1 (8.3%)	5 (41.7%)	5 (41.7%)	1 (8.3%)	12
Total	4	7	17	10	38

Table 10 Maturity levels by FRS family

Finally, we are able to compare the respondent's self-assessment of organisational maturity with the approach outlined above. Grouping the ten-point scale to create a similar 4-layer model, we find the two methods agree, or are no more than one class apart, in around 80% of cases.

A word of caution is in order here. The crude big data maturity measure we have extracted from the survey questions is very much a rough first approximation. In particular, it does not adequately address the key issues of the repeatability of procedures, given our focus on the acquisition, analysis and use of just one data set, and only touches on the question of improvability or optimisation of processes. Fuller inclusion of these criteria may paint a different picture of the big data maturity of the FRS. Nevertheless, the crude model here might serve to stimulate some further development in this direction.

Recommendations and Next Steps

The picture that emerges from this research is of a relatively skilled and confident, if also fragmented and perhaps under-resourced, data management and analytics community within the FRS. Although FRS vary in their degree of maturity, the majority have managed to negotiate, or have plans to negotiate, the often daunting journey from data to decision-making using the big data provided by the SAfER project, together with other sources, and many have a strong appetite for further data.

1. Centralised Data Cleansing and Preparation

The largest single point of frustration for most analysts has been the lack of high resolution geocoding – specifically a UPRN – in data shared by other public services.

In the case of the Exeter Data, this reflects the very different conceptual frames that operate between the FRS and the health service. The FRS have traditionally had a strongly spatial focus on specific locations – households and other properties – that are then qualified in terms of the kind of people that inhabit them. The health service, by contrast, is focused on individuals and individual records, qualified by their place of residence.

While these may appear to be equivalent, they imply very different data structures and data quality standards. While there is increasing use of GIS systems in health, administrative health data has not traditionally been effectively geocoded: patients belong to a practice, which is grouped into a CCG area. Patients have an address, but the purpose of this address is more for communication and checking identity than for service delivery.

To translate from the person-focus of health to the property-focus of the FRS requires significant work. The lack of a UPRN was the most visible and, according to our respondents, the most significant aspect of this work. The prospect of having to undertake the task of creating a UPRN for each record is, in some cases, enough to deter analysts from working with big data sets.

To the extent that it is possible, it would be highly beneficial to undertake preparation and cleansing, and in particular to add a UPRN to records, centrally in future big data initiatives.

There are a number of constraints on being able to do this effectively and it is unlikely that central processing very large data sets created for other purposes will produce a completely “clean” data set – large numbers of addresses would remain un-matched. However, even if only 85-90% of addresses were automatically matched centrally, this would make the remaining data cleansing significantly easier for individual analysts and would enable some of the geographical misallocation of records to be more easily addressed.

2. Closing the Feedback Loop: Improved Evaluation of Big Data

Assessing the value of big data sets, such as the Exeter Data, was mentioned by some, but by no means all, respondents. Some evaluation of the real value of the data – whether the data is helping crews to target a larger proportion of appropriate households for Safe and Well visits and other prevention activities, and avoiding wasting time on lower risk households – is essential if tools and techniques are to be refined and developed.

At the most basic level, only a few of our respondents were aware of attempts to gather feedback from the activity of crews which were being directed by analysis of Exeter (and other) Data, although where this had happened there were reports that crews felt that they were being directed to use their time more effectively.

Over the longer term, some analysis of the value of improved targeting of prevention activities on the number and severity of incidents is also needed. With regards to formal Risk Stratification, the main kind of modelling used with the Exeter Data, the importance of have an external check has been widely recognised. For example, NHS England have argued that

> The data generated in any risk stratification programme should be used in a feedback loop to improve the performance of the programme.[14]

However, such risk stratification evaluation is complex and can be hampered the problem of small numbers of incidents. FRS should consider working collaboratively with each other and pool their data for this kind of analysis.

There is a good case for an organisation providing central co-ordination and guidelines for the effective and practical evaluation of data-driven preventative work, both in terms of short-term operational efficiency and in terms of longer-term outcomes.

3. Supporting the Work of Data Analysts

Our respondents were mainly the data analysts who had worked directly with the Exeter Data. They form a generally confident and engaged set of individuals who rate their own data management and data analysis skills quite highly. Where they have struggled, they have often reported a lack of resource – generally their own time – rather than a lack of skills or software. Nevertheless, many respondents were keen to work with, and learn from, other data analysts, in particular from other FRS.

Tools and technique in the field of big data or predictive analytics are developing rapidly and keeping up with the cutting edge may be a challenge. Much of the development in this area has been driven by commercial marketing businesses (for example the ACORN or MOSAIC classification), so there is often some 'translation' work required to enable them to serve public service goals.

[14] NHS England (2015) *Next Steps for Risk Stratification in the NHS*, 21 January, p. 6

The availability of public service data is improving. The analysts we spoke to also need to understand the possibilities and limitations of the increasingly, if patchily, available data from other public services. This will require considerable learning, and some FRS have important experience which they can share.

Finally, while formal data management skills were generally deemed adequate, there was less confidence around the negotiation of access to data in local authorities and other partners. This was often seen as particularly challenging and some support, possibly in conjunction with the Centre of Excellence in Information Sharing, could be productive.

There is, therefore, a good case for CFOA, with others, to lead the creation of a suitable networking framework for this increasingly important group within the FRS.

4. Tactical Learning and promising practices

Our respondents provided a lot of small, quite focused tactical tips, tricks and insights that need to be more widely shared. We give four examples below:

Consider using tranches of data first

Some respondents talked about creating sub sets of data so that they could focus on defined geographical areas, mostly cities but not always. The benefit of this is that it provides quick wins early on. By choosing a small area with pre-existing datasets that support risk identification, the Exeter Data can quickly be used to refine and target. Once the approach is proven, it can be scaled up. It is less onerous and daunting to take a chunk of a big data set and test it before applying approaches to tens, sometimes hundreds of thousands of records and wondering if it's the right way forward.

Not all properties are equal

We found that many fire and rescue services were removing properties from the data set where they were clearly not people's homes. Looking for clusters of postcodes was a common approach. This leads quickly to the identification of care homes, which while they are homes of older people, have different fire safety legislation that applies and therefore tend to be outside of the usual home fire safety visit approach. Hospitals were also removed. Conversely, sheltered housing was not so easy to locate. Clusters also reveal sheltered housing complexes which have often already been identified through other data sets and therefore do not provide new insight, simply confirming existing knowledge.

The hidden householder

While the Exeter Data includes details of people over 65 it does not reveal their living status. It may reveal couples who are both over 65 and live at the same address and these duplicate records can be aggregated together for the purposes of targeting. But, what the data set does not do is identify where those with single address records are living with someone under 65. It is hard to know what to do about this as there is no obvious data set to add to this to provide additional intelligence.

The key to unlocking data

The common response across this study was that the Exeter Data is not useful on its own. Its power comes from layering it with other data sets to hone targeting activity. Many respondents had gained access to other data sets from local partners. However, many also said that they were struggling to get past the trust issues that arose from partners who did not want to, or could not see the benefit of sharing data. The key for many was to enlist the help of strategic managers higher up the organisation to have a conversation with their peer in the partner organisation. Only by building relationships and trust could additional data sets be unlocked.

5. Further Research

The **development of a comprehensive 'Maturity Capability Model,'** drawing on models developed elsewhere and more intensive research on a small number of leading FRS, could create a useful tool for FRS to benchmark themselves.

A fully developed model would need to take account of the perspectives of the analysts interviewed in this study, but also senior leadership and policy teams and operational managers within the FRS. The IBM maturity model for big data and analytics,[15] for example, addresses six issues - business strategy; information management; analytics; culture and operational execution; architecture; governance. It classifies each of these issues on a five-point scale from 'ad hoc' to 'breakaway.'

Progress towards becoming a fully data-led organisation represents a complex organisational change, requiring effective leadership and culture change as well as operational and policy innovation. Effective use of data requires the development of new attitudes towards and perspectives on data and information.

Sharing data with other public service organisations requires the development of insight and understanding as well as diplomatic and negotiating skills. Effective data stewardship implies the development of practical and ethical skills as well as technical skills. Data-driven organisational flexibility requires new relations of trust between information managers and analysts and senior leadership roles. Further, in-depth **research into the questions of organisational culture and organisational leadership would therefore provide valuable lessons.**

[15] http://www.ibmbigdatahub.com/blog/maturity-model-big-data-and-analytics

Appendices

1. References
2. Questionnaire
3. Key Informant Topic Guide
4. Data “Wish list”
5. Case studies

Appendix 1: References

Andrews, R. (2010) The impact of modernisation on fire authority performance: an empirical evaluation, *Policy and Politics,* 38(4): 599-617.

Arch, B.N. and Thurston, M.N. (2013) An assessment of the impact of home safety assessments on fires and fire-related injuries: a case study of Cheshire Fire and Rescue Service, *Journal of Public Health*,35(2): 200-205.

Clarke, A. and Miles, J.C. (2012) Strategic Fire and Rescue Service decision making using evolutionary algorithms, *Advances in Engineering Software*, 50: 29–36.

Hastie, C. and Searle, R. (2016) Socio-economic and demographic predictors of accidental dwelling fire rates, *Fire Safety Journal*, 84: 50–56.

Higgins, E., Taylor, M. and Francis, H. (2012) A Systemic Approach to Fire Prevention Support, *Systemic Practice and Action Research*, 25:393–406.

Higgins, E., Taylor, M., Jones, M. and Lisboa, P.J.G. (2013) Understanding community fire risk-A spatial model for targeting fire prevention activities, *Fire Safety Journal,* 62: SI20-SI29.

Jennings, C. R. (2013) Social and economic characteristics as determinants of residential fire risk in urban neighbourhoods: A review of the literature, *Fire Safety Journal* 62: 13–19.

Matheson, K., Manning, R. and Williams, S. (2011) From Brigade to Service: An Examination of the Role of Fire and Rescue Services in Modern Local Government, *Local Government Studies*, 37(4): 451-465.

Murphy, P. and Greenhalgh, K. (2013) Performance management in fire and rescue services, *Public Money & Management*, 33(3): 225-232.

NHS England (2015) *Next Steps for Risk Stratification in the NHS*, 21 January.

O'Grady, N. (2015) Data, interface, security: Assembling technologies that govern the future, *Geoforum*, 64: 130-137.

ONS (2105) Population estimates of those aged 16 to 64; 65 and over; 85 and over; for local areas, UK, 1992 to 2014, Reference 004835 [https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/adhocs/004835populationestimatesofthoseaged16to6465andover85andoverforlocalareasuk1992to2014

Paulk, M.C., Curtis B., Chrissis, M. B. and Weber, C.V. (1993) *Capability Maturity Model for Software, Version 1.1,* Technical report to the Joint Programme Office of the US Department of Defense.

Taylor, M. J., Higgins, E. and Francis, H. (2015) A Systemic Approach to Multi-Agency Community Safety, *Systems Research and Behavioral Science,* 32(3): 344-357.

Taylor, M., Higgins, E., Lisboa, P., Jarman, I. and Hussain, A. (2016) Community fire prevention via population segmentation modelling, *Community Development Journal,* 51(2): 229–247.

Taylor, M.J., Higgins, E., Francis, M. and Lisboa, P. (2011) Managing unintentional dwelling fire risk, *Journal of Risk Research*, 14(10): 1207-1218.

Appendix 2 Questionnaire

CFOA Baseline Study of the take-up and use of Big Data in FRS in England

FRS Identifier	

A: About the Informant

A.1 What is your name (or confirm name)?

A.2 What is your role (Job Title) (or confirm job title)?

A.3 How many years have you worked in the FRS

<5		5-10		11-20		>20

A.4 Please give a self-assessment of your information management - data handling and governance - skills (i.e., on scale of 1-10, where 10 is a world class expert, how would you rate your information management and analysis skills?)

1	2	3	4	5	6	7	8	9	10

A.5 Please give a self-assessment of your analytics - statistical analysis and modelling - skills (i.e., on scale of 1-10, where 10 is a world class expert, how would you rate your information management and analysis skills?)

1	2	3	4	5	6	7	8	9	10

A.6 Please give a self-assessment of your organisation's maturity in working with large data sets (i.e., on scale of 1-10, where 10 is a world-leading, how would you rate your organisation?)

1	2	3	4	5	6	7	8	9	10

B. Acquisition and Storage

B.0 How many Exeter Records did you receive?

B.1 How has the Exeter Data been stored in your organisation? Please tick all that apply.

Civica CFRMIS	Other risk management software	Local server PC or lap top	Cloud storage	Removable storage

Other (please specify):

B.2 Who, in your organisation, has access to the Exeter Data? (Tick all that apply)

Chief Fire Office	Senior Management Board members	Research and Analysis specialists	External partners (e.g, Universities or Consultants)
Other (please specify):			

B.3 Have you seen the Information Sharing Agreement that covers the Exeter data set?

Yes	No	Don't Know

B.4 How is access to the Exeter data controlled in practice? [Free text. Prompt for: provisions of data sharing agreement; password protected file; role based access policy; data security policy; physical media are signed out]

B.5 What challenges, if any, did the acquisition and storage of the Exeter Data provide for your organisation?

Lack of necessary resources	
Lack of data management skills	
Lack of data analysis skills	
Lack of technical infrastructure	
Other (please specify)	

B.6 How have you overcome, or attempted to overcome, these challenges?

C. Data Preparation and Cleansing

C.1 Has your organisation sought to review, document, prepare and cleanse the Exeter data?

Yes	No	Don't Know

C.2 Did you find any of the following issues with the data?

Spelling mistakes/typos	
Implausible dates of birth	
Postcodes from outside your area	
Missing postcodes	
Lack of UPRN	
Other (Please specify)	

C.3 What challenges have data preparation and cleansing presented for your organisation? [If necessary, prompt for lack of necessary resource, skills shortages, lack of technical infrastructure, etc.]

Time to carry out the work	
strategies for filling the gaps in the data	
the limitations of the ISA;	
reconciling address data	
Other (please specify)	

C.4 How has your organisation overcome, or attempted to overcome, these challenges?

D. Data Linkage, Layering and Integration

D.1 Has your organisation linked the Exeter data to, or layered it with, existing data sets or products?

Yes	No	Don't Know

D.2 If so, which data sets or products? (please tick all that apply)

Experian MOSAIC	
Caci ACORN	
IRS/Incident recording system	
Other GIS (Geographical Information System)	
Census Data	
NLPG/Geoplace Data	
Other (please specify)	

D.3 Does your organisation *plan* to link the Exeter data to other (further) data sets or products and if so which ones?

Yes	No	Don't Know

D.4 If so, which data sets or products? Tick all that apply.

Experian MOSAIC	
Caci ACORN	
IRS/Incident recording system	
Other GIS (Geographical Information System)	
Census Data	
NLPG/Geoplace Data	
Other (please specify)	

D.5 What challenges have data linkage and integration presented for your organisation? Tick all that apply.

Lack of necessary resources including time	
Lack of data management skills	
Lack of data analysis skills	
Lack of technical infrastructure	
Other (please specify)	

D.6 How has your organisation overcome, or attempted to overcome, these challenges?

E. Data Analysis, Modelling and Visualisation

E.1 Has your organisation undertaken any analysis, modelling or visualisation using the Exeter data (either on its own or linked other data sets)?

Yes	No	Don’t Know

E.2 If so, please give details?

E.6 What challenges have data analysis, modelling and visualisation presented for your organisation?

Lack of necessary resources (time)	
Lack of data management skills	
Lack of data analysis skills	
Lack of technical infrastructure (hardware and software)	
Other (please specify)	

E.7 How has your organisation overcome, or attempted to overcome, these challenges?

E.3 Does your organisation *plan* to undertaken any (further) analysis, modelling or visualisation using the Exeter data (either on its own or linked other data sets)?

Yes	No	Don’t Know

E.4 If so, please give details?

E.5 Have you worked with, or do you plan to work with, other organisations to analyse the Exeter data?

Other FRS	
Consultants	
Software vendors	
Universities	
Other (please specify)	

F. Use of the Exeter Data in Decision-making (planning, targeting, operational decision-making)

F.1 Has your organisation made use of analysis, models or visualisation that draw on the Exeter data to support strategic or operational decisions?

Yes	No	Don’t Know

F.2 If yes, were these in the following areas?

Improved targeting	
Improved risk assessment/profiling	
Improved resource planning and allocation	
Other (please specify)	

F.3 Do you have or expect to have evidence of the impact of these decisions on FRS outcomes or KPIs?

Yes	No	Don’t Know

F.4 If yes, please give details.

F.7 What challenges has using, or planning to use, analyses, models or visualisation drawing on the Exeter data to support decision making presented for your organisation?

Lack of necessary resources (time)	
Lack of data management skills	
Lack of data analysis skills	
Lack of technical infrastructure (hardware and software)	
Other (please specify)	

F.8 How has your organisation overcome, or attempted to overcome, these challenges?

F.5 Does your organisation *plan* to make use of analyses, models or visualisation of the Exeter data to support strategic or operational decision-making?

Yes	No	Don't Know

F.6 If yes, please give details.

F.9 What challenges *do you envisage* in using analyses, models or visualisation drawing on the Exeter data to support decision making presented for your organisation?

Lack of necessary resources (time)	
Lack of data management skills	
Lack of data analysis skills	
Lack of technical infrastructure (hardware and software)	
Other (please specify)	

G. Evaluation and development

G.1 How does your organisation plan to evaluate the benefits of the Exeter data?

G.2 Would any of the following forms of support help your organisation to get the most out large data sets such as the Exeter data?

Training in data management and/or analysis techniques	
Specialist software for data management and analysis	
Expert guidance on the use of data in decision making from outside the FRS	
Mentoring from other Fire and Rescue Services	
Other (please specify)	

H. Final Questions

H.1 Thinking about **all of the data** that your organisation uses to plan, deliver and evaluate fire and rescue services, how "mature" would you rate **your organisation** to be on scale of 1-10, where 10 is a world leading user of data?

1	2	3	4	5	6	7	8	9	10

H.2 Are there other sources of data, to which your organisation does not currently have access, that you believe would be valuable to your organisation?

H.3 Is there any other information that you think would be valuable to CFOA concerning the use of large data sets such as the Exeter data by your organisation?

H.4 Would you be interested in being approached to take part in further research on the use of data and analytics in the FRS?

Yes	No	Not Sure

Thank You. We will contact you shortly with an anonymised summary of the results of the survey.

Appendix 3 Key Informant Topic Guide

CFOA Baseline Study of the take-up and use of Big Data in FRS in England

Key informant context interview Topic Guide v.01

Preface: [As per survey questions]

1. Could you tell me something about your current role in the Fire and Rescue Service?

2. Could you tell something about your background and experience? [Prompt for education, skills, qualifications, experience in previous roles and posts]

3. Could you briefly tell me how you came to be involved with the sharing of the Exeter data with the Fire and Rescue Service.

4. What, from your point of view, were the main expected benefits for the Fire and Rescue Services, individually and collectively, of acquiring and using the Exeter Data?

5. What, from your point of view, have been the main challenges confronted in acquiring and using the Exeter Data? [prompt for practical, technical, organisational, legal, governance, political, ethical]

6. What from your point of view have been the main lessons that have been learned from the sharing of the Exeter data with Fire and Rescue Services? [prompt for practical, technical, organisational, legal, governance, political, ethical]

7. How do you envisage the future of (big) data sharing between the FRS and other public services?

8. What, if anything, could CFOA do to improve the effectiveness of data sharing for the Fire and Rescue Service?

9. Are there any issues that we haven't covered that you would like to raise?

10. Would you be happy to be involved in future research on the use of data in the Fire and Rescue Service?

Thank You. We will send you, in due course, a record of this interview to check that we have not misunderstood or misrepresented you. We will also send you a summary of the whole research project when it is complete.

Appendix 4 “Wish list” of Data with requesting FRS

Data Set Requested	Requesting FRS
A&E Burns admissions	London; Staffordshire
Adult Social Care Data	Leicestershire; West Sussex
Age UK Loneliness dataset	Oxfordshire
Alcohol abuse	Humberside; Cambridgeshire; Warwickshire; West Yorkshire; Hertfordshire
Ambulance data at household level	Leicestershire; Royal Berkshire
Carers Allowance	London
Clinical Commissioning Groups	Lancashire; South Yorkshire
Council Tax single occupancy data	London; Humberside; Durham & Darlington
Dementia	Cleveland
Disability Data	Humberside; Cornwall; South Yorkshire; Oxfordshire; Shropshire
Domestic Violence	Cornwall
Drug Abuse	Cleveland
DVLA- Ill health Data	Dorset and Wiltshire
DWP attendance allowance	Avon; Humberside; Greater Manchester
DWP benefits data	Tyne and Wear
DWP mobility data	Dorset & Wiltshire; Oxfordshire
Environmental Health	Devon and Somerset
Families with Young Children	Shropshire
Fires that go unreported	Essex
Flooding	Cornwall
Food standards data	Devon and Somerset
Frailty index from NHS	Derbyshire; Kent
Frequent 999 callers	Cornwall; Cleveland
GP 'Frequent Flyers'	Lancashire; Warwickshire; Cornwall; Staffordshire; Kent; West Yorkshire; Hertfordshire
Health Data	Greater Manchester
Hoarding	Avon
Housing Association Data	Staffordshire
Living Alone	Staffordshire
Long Term Conditions	Durham & Darlington; Staffordshire
Mental Health	Humberside; Cambridgeshire; South Yorkshire; Cleveland; Hertfordshire
NHS Number	Lancashire
Off Grid Power Data	Derbyshire
Oxgen Suppliers	Avon
Police data for ASB	Warwickshire; Essex; Royal Berkshire; Tyne and Wear
Private Care providers	Essex
Public Health Data	Buckinghamshire; Warwickshire
Registered Disability	London
Rented or social housing tenants	Humberside; Durham & Darlington; Cornwall
Road Traffic Collision data	Royal Berkshire
Single parent families	Durham and Darlington
Smoking	Humberside; Warwickshire; Cornwall; South Yorkshire
Social Care Data	Devon & Somerset; Derbyshire; West Sussex; East Sussex
Social services data on vulnerable families	Essex; West Yorkshire
Valuation Office Data	London
Victims of Fraud/Scams	East Sussex
Vulnerable People	Devon & Somerset

Appendix 5a Dorset and Wiltshire Case Study

Dorset and Wiltshire case study

Dan Cleaver is new to the fire and rescue service. He joined Dorset Fire and Rescue Service two years ago and is now Community & Partnership Data Co-ordinator for Dorset and Wiltshire Fire and Rescue Service.

Before the 2016 merger, Dorset Fire and Rescue Service received 191,286 Exeter data records. On examining and cleansing the data, they discovered 1059 records with no date of birth and 14,068 records with addresses outside of Dorset.

Data were geocoded using Bluelight Symphony iMatch (http://www.aligned-assets.co.uk/products/imatch/) to apply a UPRN. Around 80-85% of records were successfully matched to a UPRN by the software so a significant volume of manual cleansing was still required, as Dan explains.

"At this stage, we knew not all the records would automatically be matched with a UPRN after putting the file through iMatch, but we were unsure how many would be unmatched and why. We had already removed commercial properties, care homes, residential homes and hospitals. We removed them because we were going to use the Exeter Data to target properties for a home safety check."

The scale of the data cleansing task is significant and while Dan's experience isn't unique, his technical approach is interesting to capture. He shared his process.

"The next stage was the most time consuming. We found common issues such as spelling mistakes and addresses outside of Dorset that we hadn't picked up on because the county name was absent in the address line. We also found incorrect formatting of postcodes.

"The different issues by themselves are insignificant but when combined, they amount to a significant amount of data to cleanse, to be able to match it to a UPRN. It then requires a semi-manual intervention using Excel formulas to reformat the addresses to be able to put through iMatch."

Since the merger, Dan has the task of making sense of the Wiltshire Exeter data records as well; these had not been analysed previously. So he had to add a further 140,000 records and he now has nearly half a million records to manage when including other partner data.

The data was analysed in Pinpoint, a GIS interface to ARCInfo. The Exeter Data was added as an additional layer. Dan explained how he did this.

"With Pinpoint, the way we have used Exeter Data is to integrate it with our internal data and external partner data sets. We have daily updates of our response data as well as our incident data and home safety checks. From this we can categorise household risk using high, medium or low; this can of course change over time based on the data.

"This is then integrated with other external partner data sets such as council tax single person discount data and higher rate disability benefit data. The partner targeting approach allows us to select a household within the high-risk category and then choose to further drill down to select only households that have a single person occupying it who is over 65 and is outside of our response standards. We can then target the most at risk and vulnerable households within different areas."

Dorset and Wiltshire have further plans to split the data down further including separating male and female and splitting the age groups into "65-84" and "over 85." This will enable risk stratification to pick up on high risk individuals over 85 but also the males in the 65-84 age range living alone who are also at higher risk. With more detailed data, models can be created more flexibly and tailored for local areas.

"It gives us a bit more flexibility in honing our targeting for fire stations in different areas, as each area has a different risk associated with them. It gives the user more freedom to select what is appropriate for their area. Whether this becomes a success is something we will find out."

Dorset and Wiltshire are also trying to move from mapping the existing distribution of risk towards predicting how that risk profile or risk landscape is changing.

"Something we are looking to do is to be able to get better data that helps us to better predict when someone is on the path to becoming vulnerable or at risk before they actually are. Different organisations are already starting to conduct research into things like loneliness and showing the link from loneliness leading onto other health related issues and this is something that could help us better identify at risk and vulnerable people."

Dan was clear that there is considerable further potential in the Exeter Data. "We're probably like a lot of fire services, still trying to understand how to use it best."

Indeed he is right. The research undertaken here has revealed that there is a lot of interesting work going on with the Exeter Data. There is a lot of repetition and learning that could be easily shared to reduce the time being taken to make sense of it and use the data to increase local understanding of risk.

Appendix 5b East Sussex Fire and Rescue Service

Chris Fry is Community Risk Analyst with East Sussex Fire and Rescue Service (ESFRS). He has been there for 12 years and works with just one other analyst.

ESFRS has developed a risk tool called the Cube, which is their way of identifying vulnerable people and vulnerable areas. Previously they used FSEC (Fire Services Emergency Cover) to locate risk but found they were having fire fatalities in low risk areas according to FSEC, so they wanted to become more granular with their analysis and that was what led them to use MOSAIC.

In the Cube, they have MOSAIC lifestyle data to identify which people have fires or are more likely to do so. For every household they are looking to identify injury propensity and the way in which they are most likely to want to be contacted for safety information. In addition to this, they look at how far away someone lives from a fire station and then they categorise their 'rurality' risk. They also use deprivation data from IMD (Index of Multiple Deprivation). FSEC risk is still in there, providing the likelihood of dying. HFSV data is included as it shows when someone had received a visit and the data they gathered about that household at the time.

Chris has been able to add adult social care data to the Cube as well. This dataset includes a risk rating for each person as determined by their own formula. ESFRS has a data sharing agreement with the county of East Sussex for this dataset of 14,000 addresses. By adding in the Exeter data, ESFRS is able to focus in on age and another agency's understanding of risk.

ESFRS carry out 10,000 Home Fire Safety Visits a year and have a commitment to visit those identified in the adult social care dataset, so they are already committed to 18 months work on just this. "Obviously the Exeter dataset is enormous", says Chris. They have 51,000 records of people over 80 in their area.

Layering in the Exeter dataset allows ESFRS to create subsets of adult social care data and effectively apply an age filter which they had not been able to do before. This means that the commitment to one data provider is further enhanced not only by Exeter but by all the other layers of data so that they really focus on the most vulnerable.

Interestingly, by doing this they have already identified householders who according to adult social care data are in a low risk group, but are from a fire point of view at higher risk due to age and lack of previous exposure to the fire and rescue service via its HFSV programme and therefore need to be prioritised for visits. It's Exeter data that is allowing that additional refinement to take place.

It is an example of how one organisation's view of risk can be quite different to another's and that the multiple layering of datasets can lead to a more sophisticated and nuanced view of vulnerability.

Given this commitment, it is good to see that ESFRS continues to seek additional datasets to add to its Cube to continue to refine risk. They are looking at data from trading standards showing people who have been victims of scams and frauds. They are also at an early stage in negotiating access to frailty data from

GPs, particularly those who have a history of falls or recent access to healthcare systems.

Chris talks about improving the technical architecture both in terms of robustness of the tools to deal with growing amounts of data (he currently uses Excel) and also the desire to allow others to access it through an intuitive front end. All of which takes time and cost money. “At the moment, it’s just me. There’s potential for it to develop.”

Chris, like many of the people interviewed for this research, has a deep commitment to improving the work he and his fire and rescue service do in terms of refining risk and identifying those who benefit most from interventions like HFSVs and Safe and Well Visits. Despite a simple technical infrastructure, Chris and many like him in other fire and rescue services are making the most of the tools at their disposal to make a real difference, one dataset at a time.

Appendix 5c Cleveland Fire Brigade

Much of the experience of Tim Graham, Head of Risk and Performance at Cleveland Fire Brigade is similar to others interviewed for this study. His service received 135,000 records, but he found he had 30,000 records that were from outside his area by using his GIS software and local knowledge of postcodes. He dealt with the UPRN issues in a similar way to others.

More fundamentally for Cleveland, the addition of the Exeter dataset has led to a change of policy. Previously HFSVs were carried out based on either a response to a request or through ‘cold calling’. With the refinement of knowing the age of householders, Cleveland has moved away from that approach and is generating its 15,000 HFSVs with age as a proxy for risk and only visiting those over 65. This isn’t unique to Cleveland, during the research it became clear that other fire and rescue services were using data to move from a responsive to a more proactive approach to targeting risk.

However, Tim was the only respondent to mention the electoral roll and how important he thinks it is for fire crews to have the names of householders when they knock on the door. In moving from the reactive to the proactive approach, Cleveland fears losing the personal touch: they don’t know who they are visiting now. They only have ‘the occupier’ for the address and, Tim says, “that the absence of a name proves quite difficult for getting into peoples’ houses.”

Adding in the name from the electoral roll is one way of overcoming this, and while not a panacea, it’s a good step forward in improving the personal level of the intervention.

Everyone interviewed for this research was asked to provide a wish list of datasets they would like to have access to in order to refine their risk stratification. Tim was the only one to refer to emergency medical response activity – or co-responding. He says that while the fire and rescue service attends and intervenes, it has no record of the outcomes once the patient transfers to the ambulance service. “We never know what happens to the patient.”

For him, this is a lost opportunity; to trace a patient through the healthcare system and know whether they would have benefited from a preventative intervention. It’s an interesting idea and certainly linking hospital discharge with the fire prevention work of fire and rescue services is not new, it’s just not routine business across the country.

Talking to 35 different analysts across the fire and rescue service in England revealed a community that was disparate and only loosely joined up. Some services, like Tim in Cleveland, talked about their family group – organisations that are similar to them – as a community in which to discuss the issues he has faced with this new dataset. “What I’m intending to do is see how they (the family group) is using the Exeter data. Is there anything I can pick up from them or they can learn from me on how we actually use it.”

Tim hasn’t had that conversation yet with his family group but thinks that a year on, maybe 15 months in, this will be incredibly helpful. It is the tactical level of detail that he is after and like many others, he would benefit greatly from the reassurance and the learning that comes from meeting up with those with a shared experience.

One of the recommendations from this research is that a community of analysts should be supported and developed. Whether that is a virtual one or a face to face one, either way, the opportunity to share experience and swap tactics and solutions must be beneficial to reduce ‘reinventing the wheel’ and opening up more time to really get to grips with this and other new datasets in the future.

Appendix 5d Warwickshire Fire and Rescue Service

Vanessa Belton is the Performance and Improvement Business Partner. She works for Warwickshire County Council where fire and rescue is a department, spending half of her time on fire and half on communities work. Having worked around fire and rescue data since 2001, Vanessa is experienced and knowledgeable about this area.

As with many other staff interviewed for this research, Vanessa spent a lot of time working out how to get the dataset into a usable form. She says, “We spent quite a lot of time on that one because the data itself was a little messy. We had lots of everybody else's [data]. We weren't sure if we got a complete set and I'm still not convinced that we have. So we did spend quite a lot of time making sure that it was as accurate as it could be and it was in a usable format.”

Once the dataset is cleansed, the next stage is to work out how to link it with other datasets. Many respondents said that the dataset by itself was of little value but layered with others it became very powerful. In the case of Warwickshire, Vanessa says “We scratched our heads and said ‘how would we best use this?’ Because we know it’s an absolutely fantastic source of information, exactly the right people we want to target. It’s about how we use it to its best potential really.”

Vanessa is enthusiastic about the acquisition of this large dataset. She has relished the challenge of getting her teeth into it and using it to hone in on targeting vulnerability in her area. “It’s a nice thing to do” she concludes “because you know it’s going to have a benefit at the end of the day. I’ve enjoyed doing it really.”

Warwickshire has a lot of retained duty stations and using the Exeter dataset Vanessa has identified vulnerable people in areas that are served by these stations. “It shows in particular areas that we actually need to target RDS stations because an awful lot of vulnerable people are there and we haven’t necessarily put as much effort into those areas in the past and this has shown this significantly.”

During the research, it became clear that there was little evaluation work taking place in fire and rescue services which looked specifically at the impact of the Exeter dataset. It is likely that this is because it is early days and some fire and rescue services have only had the dataset in a usable format for a few months.

For others, like Warwickshire, there is some anecdotal evidence that suggests the use of the dataset is improving targeting (and by extension, would hopefully improve outcomes, although there is no evidence to support this at this stage).

“The feedback is that they [the fire crews] are going to what they perceive to be exactly the right people. They’re not knocking on the door and then somebody completely random is opening it. The crews are really happy with it. It’s definitely the way that we’ll go in the future with some targeting. “

It is a good start for Warwickshire and while much of what Vanessa has experienced is not unique, her journey with the data demonstrates how quickly the acquisition of a new dataset can make a difference to targeting risk. With more time, experience and a refresh of the data, Vanessa and Warwickshire can really start to make a difference in finding those who most need the help and advice in her county and potentially improve outcomes overall.